Review of the Scientific program and information relevant to the International Dolphin Conservation Program Act (IDCPA) held at the Southwest Fisheries Science Centre, La Jolla, California, 8 to 11 March, 1999.

By

Stephen J. Smith
Head, Molluscan Fisheries Section
Department of Fisheries and Oceans
Dartmouth, Canada

The terms of reference for this review which operated under the University of Miami's Centre of Independent Experts were as follows:

1. To evaluate if the mail reviews of the scientific program were appropriately taken into account (i.e., comments incorporated or actually rebutted), and

2. To evaluate if the final conclusions in the draft report to congress were supported by the research to date.

Given that my expertise mainly concerns stock assessment models and sampling theory, I will concentrate on those aspects of the program. The other two members of the review team will deal with the habitat and stress issues.

1. Mail Reviews

Reviews by Marine Mammal Commission (MMC)

Comments by the MMC were contained in two letters dated 8 January and 12 February 1999. In the first letter, the following points were raised concerning the stock assessment model or the research survey.

1. What value would be used for the estimated maximum growth rate (r_{max})? Letter noted that values of 3 and 4 percent were discussed by NMFS but no consensus reached.

NMFS responded by stating that the values used in the analyses of 1.5 and 1.7 percent for northeastern spotted and eastern spinner dolphin populations, respectively, were based on calculations from the data for 1975 to 1991. While, the idea of incorporating a default value of 3 or 4 percent would be worth investigating, the current approach of using an estimated value will be retained.

2. NMFS should consider publishing the details of the proposed decision analysis framework.

Document giving details of the framework prepared by Prof. D. Goodman (Montana State) and sent to MMC (received 1 February, 1999; see section on letter from MMC dated 12 February below).

3. Due to the lack of historical information on coastal spotted dolphin, NMFS should consider an alternative approach be used for making a determination for this stock. It was further recommended that NMFS obtain and review information on the frequency at which sets are made on this stock and the numbers of coastal spotted dolphins that are chased and encircled relative to the estimated stock size.

NMFS responded that while this proposal has merit there are difficulties in obtaining or interpreting some of this information. They proposed that at least an estimate of the average annual frequency of sets on spotted dolphins within the coastal stock's range would be provided for the 2002 finding.

The comments contained in the letter of 12 February mainly concerned editorial changes to the three background documents detailing methods and results for the 1998 research survey, the literature review of the potential influence of fisheries induced stress and an preliminary investigation of dolphin habitat variability. In addition, receipt of a document on the decision analysis framework was acknowledged and the MMC reported that it was satisfied with the documentation and the framework itself.

The editorial changes suggested for the document on the 1998 research survey were all taken in account in the revised version dated 26 February 1999.

I conclude that all of the MMC's comments were either incorporated or objectively rebutted by NMFS staff.

Reviews by the external referees

Technical comments on the line transect methodology were submitted via email by Dr. K. Burnham, Dr. S. Buckland and Dr. Jeff Laake. All of these individuals have participated in a large way in developing the methodology and software for the line transect methods used to estimate abundance from the research survey and the kill estimates and population indices from the tuna vessel observer data (TVOD). Given their involvement, their reviews can hardly be construed as independent outside reviews. Overall, their comments mainly dealt with matters that could provide for future improvements to the surveys and the methods. While the comments were pertinent to the line transect methodology and the reviewers are experts in the field, future reviews should at least include outside experts in survey sampling design.

Reviews by the Inter-American Tropical Tuna Commission (IATTC)

Three letters dated 14 January, 3 February and 17 February from the IATTC to NMFS commenting on the scientific program were provided to the reviewers prior to the week of 8-11 March.

14 January 1999:

1. Concerns about the decision framework and the interpretation of the TVOD index.

3 February 1999:

1. Comments on habitat variability and stress paper.

These will be handled by Drs. Olson and St. Aubin.

17 February 1999:

1. Reiterated comments from 14 January on problems with the decision framework in response to NMFS response to the original comments by the IATTC.
2. IATTC contends that TVOD index does show that population growth rate of northeastern offshore spotted dolphin has been between 2 and 4 percent in the 1993-1998 period.
3. Problems of the comparison between TVOD and MOPS.
4. Comments on small CV's for the TVOD, especially in the most recent years.
5. Questioned usefulness of comparing population growth rates between the two periods (pre and post 1991).
6. Questioned designation of reduction in population growth rate as being due to unreported additional mortality rather than change in carrying capacity or some other answer.
7. Refers to problems observed in a "strategic" analysis currently being done by IATTC (not provided) of TVOD data.

The main issues raised by the IATTC in their letters of 14 January and 17 February concerned the decision rule defined by NMFS and the interpretation of the TVOD data. Overall, the NMFS response was to try to agree to disagree on the decision rule and to point out that the IATTC appeared to use the TVOD as a relative index of abundance so why couldn't NMFS. I discuss both these items in more detail below when discussing the model under the second term of reference. There being no middle ground on these points, it can be concluded that for the most part NMFS stuck to its point of view when drafting the Report to Congress. It was difficult to do otherwise because there was no published or widely disseminated material concerning the problems with the TVOD raised by the IATTC.

A fourth letter from the IATTC to NMFS dated 5 March was submitted to the reviewers on 8 March despite our intent to limit submissions to what could be received by the reviewers as of 2 March so that adequate time would be available to understand and evaluate the contents. NMFS staff did not have time to reply to the 5 March letter and as a result no evaluation of response on my part is possible.

2. Conclusions of Draft Report

The finding of the National Marine Fisheries Service is there is evidence that in the period since 1991, there has been for the northeastern spotted and eastern spinner dolphin populations a failure to grow at the expected rate (designated as r_{max}) of 1.5 and 1.7 percent, respectively.

The population model developed to evaluate the current status of the dolphin stocks is a Leslie matrix population model in which parameters are estimated using empirical Bayesian methods. The parameter estimates themselves are obtained as means or medians from the posterior distributions for the parameters with these distributions being generated from the input data (TVOD, research survey, TVOD estimates of dolphin kills by tuna fishing, Leslie Matrix structure with respect to age composition, selectivity, survival and fecundity) and prior distributions for the parameters of interest using a computer-intensive resampling approach (SIR method). While this model is complex in structure its behaviour with respect to the main data sets used in it can be characterised as one where the TVOD data generally determine the trend in the population estimates while the research survey generally determine the scale (parameter a). That is, the research survey data contribute more toward determining the absolute population size of the dolphin populations than trends in population size over time. Trend and scale are modified somewhat by the age composition, selectivity and assumptions in the model concerning the fecundity and age of sexual maturity. However, the main signals for the population model come from the TVOD and research survey data.

Major issues of the model:

Research survey data:

The research survey data used in the population model come from three separate programs: 1) Research vessel surveys, 1979–80, 1982–83; 2) Monitoring of porpoise stocks (MOPS), 1986–1990; 3) Stenella population abundance monitoring (SPAM), 1998 [first of three mandated surveys]. All of the survey data have been analyzed in a similar way for inclusion in the model. The general trends over the three series are highly variable for the northeastern spotted and eastern spinner dolphin populations. Both series exhibit large increases/decreases in population size that would appear to be biologically unrealistic given the expected growth rates for these populations. Additionally, all of the abundance estimates from these surveys have large coefficients of variation (CV) with most being in the range of 22 to 46 percent and 34 to 76 percent for northeastern spotted and eastern spinner dolphin populations, respectively. The 1983 estimates for both species and the 1980 estimate for the eastern spinner dolphin have very large CV's ranging from 161 to 359 percent.

As has been already noted by the reviewers and the authors of the reports, the larger CV's associated with the research survey compared with the TVOD data results in the survey data contributing more to scaling the population size estimates than in estimating the trend of the population. The only way that the research survey estimates will contribute more to the trend aspect of the model is have much lower CV's on the 1999 and 2000 surveys relative to the TVOD. While the CV's for the 1998 estimates are in the lower range of those observed for all three surveys, the 1998 levels of precision were obtained using the total sampling effort of three vessels. It appears that only two vessels will be available for 1999 and 2000 which suggests that the CV's for estimates in those years may be expected to be higher than those for 1998. Therefore, if the CV's of the estimates from the next two year's surveys end up to be no lower than previous surveys then it is unlikely that these survey estimates will in any substantial way influence the trend predicted by the model.

If sampling effort can not be increased, then it may be possible to evaluate the current survey design to see if its efficiency can be improved upon by changes in the allocation or stratification scheme. In sampling theory, stratified random survey designs result in increased precision (lower CV's) for estimates over simple random sampling design only if the design of the stratification scheme and the scheme for allocating sampling effort (days) to strata is optimal. The theory for evaluating the efficiency of a stratified random design is available for estimating means (see Smith and Gavaris 1993 for a fisheries example). The method partitions the difference between the variance for simple random sampling and the variance for a stratified random design into two components — allocation and stratification. The allocation component can be equal to zero, greater than zero or less than zero depending upon whether the number of sampling days were allocated to strata proportional to the area of the strata, proportional to the product of the strata area and strata standard deviation, or in an arbitrary manner. The strata component will always be greater than zero and its magnitude will reflect whether the within strata variances were larger than the between strata variances. If the sum of the strata and allocation terms is greater than zero then the current stratified design has resulted in a gain in precision. It is entirely possible that the strata may be meaningful but that the allocation of sets to strata is so suboptimal with respect to strata variance that the allocation term is large and less than zero resulting in a difference between the simple random sampling variance and the stratified variance that is also less than zero. This implies that the stratified design resulted in an estimate with lower precision (larger CV's) than simple random sampling. Given that allocation schemes are generally under the control of the survey program staff, changing this scheme to be optimal in some manner (as discussed above) may result in a gain of precision without additional cost. If the conditions with respect strata variability are fairly constant over time then it may be possible to use the 1998 survey to redesign the allocation scheme for the 1999 survey.

Often however, the distribution of animals is rarely constant from year to year and last year's survey may be a poor predictor of conditions in the current year. In that case, the method of adaptive allocation may more useful (Thompson and Seber 1996). In this method the total number of days for the survey d is partitioned into two groups d_1 and d_2, where d_1 is usually much larger than d_2. In the first phase of the survey d_1 days are allocated to strata either in proportion to strata area or strata variance from the previous year (if informative) and the survey is conducted. Prior to the first phase a sampling rule is set which prescribes increased sampling in a stratum if it's variance (or mean or CV, etc.) is greater than some threshold level. After completion of the first phase the d_2 days are allocated to those strata for which the threshold level was exceeded in the second phase. Thompson and Seber (1996) provide the appropriate equations to provide unbiased estimates for this kind of two phase sampling.

The situation for the line transect surveys is more complicated than the usual situation of estimating a mean. As a result explicit formulae for assessing the efficiency of the stratified design for line transect surveys do not exist. However, given the number of days observed in the 1998 survey it should be possible to assess the efficiency using sampling/Monte Carlo methods. In addition, the unbiased estimators given in Thompson and Seber (1996) may not be strictly applicable to the line transect situation, although this fact has yet to be determined. The bottom line here is that methods do exist for evaluating and improving the precision of estimates from stratified random designs in the standard situations and modifications could be possible to do the same for the line transect surveys. These approaches offer the means to improve the precision (and hence decrease the CV) for the future research surveys at no extra cost which in turn may mean that these surveys contribute more trend information to the population model than the surveys currently do.

The variances (and hence CV's) for the line transect survey are estimated using the bootstrap resampling method. Recent research in the statistical literature (Rao and Wu 1988, Sitter 1992a,b) has indicated that the application of the standard bootstrap method to situations involving complex survey designs may result in under-estimates of the true variability. Modifications have been provided by the authors cited above and these should be investigated for the line-transect survey data (see Smith 1997 for a fisheries application). These modifications may also be appropriate for the TVOD indices which also use the bootstrap to calculate variances.

TVOD data:

One of the most contentious issues raised by IATTC in its review of the population model is the use of the TVOD data as an abundance index for the dolphin populations. Note that the use of the TVOD estimates of dolphin kills in the model does not appear to be disputed by anyone.

In the most recent IATTC annual report (Anon. 1998) the TVOD indices (smoothed and including error estimates) for the northeastern stock of offshore spotted dolphins and the eastern spinner dolphins are presented as abundance indices in Figures 84 and 87, respectively. The text associated with these figures (pages 79-80) admit that there may be biases in the data and suggest "... so the resulting estimates should be treated as indices of relative abundance of the stocks, rather than estimates of their absolute abundance." The annual report uses these relative indices to comment on population trends for the various dolphin stocks affected by the tuna fishery. The NMFS population model uses the TVOD indices as relative abundance with the conversion to absolute abundance done via the parameter *a* and contributions from the research data (see above). Appendix 1 of the draft Report to Congress contains a review of the literature mainly contributed by IATTC scientists on modifications of the TVOD data so that they may be used to construct relative abundance indices.

It appears therefore that the reservations expressed by the IATTC are fairly recent and yet to be peer reviewed and published. Are the concerns of the IATTC valid? It is difficult to answer this question without a thorough analysis of the data but there appear to be a number of issues that demand some further study and explanation. The IATTC notes that the TVOD index and the total number of dolphin sets appear to track each other since at least 1981 (comparison provided by IATTC as Appendix D in their letter of 5 March, 1999). Information on trends in the number of dolphin sets is also provided from 1986 to 1996 in Figure 83 of the annual report (Anon. 1998). It should also be pointed out that the reviewers noted that a rough comparison of the TVOD with the searching model and cohort indices of abundance for yellowfin tuna (Fig. 31, Anon. 1998) appears to show coincident trends as well. One explanation for these coincident trends could be that the TVOD is simply reflecting encounter rates with dolphins but a few trends compared by eyeball hardly constitutes a definitive analysis and a more rigourous study should be made of these patterns.

The IATTC contends that there has been a change in the pattern of the data used to estimate the detection function arguing that recent changes in the addition of helicopters and bird radar may result in fewer observations near the trackline than there should be. It is interesting to note that the variability in numbers of dolphin sightings near the trackline noted for the TVOD is also evident in the 1998 survey data (Figure 5, Gerrodette document on Preliminary estimates of 1998 abundance ..., dated 26 February, 1999). The interpretation of abundance indices derived from commercial fisheries data is often complicated by technological changes to the actual process of fishing and the TVOD is probably no different in this regard. While an analysis of the TVOD to determine if technological changes to the fishing process have affected the indices is certainly warranted, the current explanation for the patterns in the data used to calculate the detection function needs more work.

The TVOD indices are derived from commercial fishing data through a process that by its nature lacks the control that one may have over a research survey. On the other hand, the TVOD has the advantage of obtaining more data in time and space than a research survey allowing it to possibly provide a more complete picture of the dynamics of the fishery and the dolphin populations. As noted above, commercial fishery based abundance indices can be influenced by changes in fishing practice and attention must be given to constantly monitoring the fishery for these changes in order to modify the indices appropriately. The use of the TVOD by the NMFS to estimate the growth rate of the dolphin populations seems to have brought the issue of evaluating the usefulness of the TVOD to a head. However, for the moment we do not have any precise answers instead we have a number of analyses that need to be done.

Definition of μ:

Prior to 1992 all fisheries-based mortality of the dolphin populations was accounted for in the population model by the reported kills, M_i from the TVOD estimates in year i. However, from 1992 to 1998 an additional parameter μ was introduced to account for additional mortality above and beyond that represented by the M_i, i.e., total fisheries-based mortality was equal set to

$M_i + \mu N_i$, where N_i is the estimated population size in year i. The quantity μ (along with the estimate of r_{max}) is used in the decision rule to assess whether there has been a failure for the population to grow at the expected rate.

This definition of additional mortality may be problematic especially if the TVOD shows an increase in dolphin abundance over the next three years. The parameter μ is defined as a constant rate of mortality over the years 1992 to 1998. In the case of the northeastern offshore spotted dolphin (Figure 2, Draft Report to Congress), 1992 was a high year for the TVOD while the index was lower and flat (no trend) from 1993 to 1998. While all of the years from 1975 to 1991 were used in the population model to estimate the population abundance in 1992, μ may be seen as being driven by one high point and five low points. If the TVOD shows an increase at a rate that may be biologically plausible over the next three years (before the 2002 finding), the five low points may still be influential enough to make the estimated μ relatively insensitive to the growth rate as indicated by the trend in the more recent years. The parameter μ may be much less sensitive if the TVOD increases by a relatively large amount in the next two years but these changes would be questioned because their apparent rate of change exceeds that expected from the dynamics of this population.

Interpretation of μ also appears to be problematic in that it can represent either unseen mortality perhaps defined in terms of deaths which occur some time after the dolphin is caught and then released from the purse seine or retardation of reproduction through either stress-related abortions of fetuses or through some stress-related inhibition of reproduction after exposure to fishing or any combination of these. Given that this mortality or decrease in reproductive output is a function of exposure to fishing activity it is difficult to understand why NMFS chose to define μ as a constant proportion of the dolphin population size rather than some function of either fishing effort or numbers of dolphins recorded as killed by fishing. As it is presently defined it is possible that μ could be estimating some level of mortality even if M_i is equal to zero. The way the model is currently constructed a value of zero for M_i could indicate that no dolphins were killed in dolphin sets or alternatively, no dolphin sets were made but dolphin sightings were still being reported by observers on tuna vessels. The utility of μ appears to be limited as currently defined and an alternative approach should be developed certainly prior to the 2002 finding.

We were provided with a fit to the data which used an exponential model for the series with a break at 1992 and no μ parameter. Indications that there had been a decrease in growth rate still remained, so μ as defined does not appear to be driving the decision but as stated above this parameter has serious limitations in its own right.

Robust to age structure:

The reviewers expressed an interest in knowing if assumptions concerning age structure and the age-specific selectivity curve affected the final results of the model with respect to detecting a failure of the dolphin populations to grow at the expected rate. Paul Wade provided us with a run of his model which included an identical selectivity for all ages and the results were similar to the run provided in the draft Report to Congress — there was still evidence for a depression in the growth rate according to the original criteria chosen. In an analysis for the same data but without age structure Wade (in press) found evidence that the dolphin population was depleted — note that the parameterization to determine depletion was different than that used for the current report but the conclusions about the status of the stock were similar.

Decision analysis Framework:

Comments forwarded by the IATTC argue that, while the decision criteria proposed (and used) may be appropriate for resource management issues, they are not appropriate for the question of product labeling, are of more legal interest than of statistical concern. I do not have the expertise to offer opinion on legal matters. However, the decision theoretic framework as a tool appears to be a sensible way of approaching this problem and the only way to deal with decision making when using a Bayesian model. The idea of a decision tree to incorporate a broad range of questions concerning possible sources for the mortality is only outlined in a cursory manner in the supporting documents. I assume that this idea will develop over time but there is not enough information to comment on at this point in time.

In conclusion then, assuming that the TVOD indices tracks dolphin population abundance and the research surveys currently index the range of population size, the population model based on the Leslie matrix with the Bayesian statistical model appears to be a useful model for evaluating growth rate changes. The use of a Bayesian model to estimate parameters leads to using a decision theoretic approach to make inferences. The main model-based weakness noted here is the definition and limitations of the parameter μ for determining if there has been a depression in the growth rate. As noted above different parametrizations for the model of the northeastern offshore spotted dolphin used in Wade (in press) and an exponential model presented to the reviewers gave indications that the stocks are still in a depressed state. The research surveys are extremely variable and suggestions have been given here for ways of evaluating and improving the current design for the next two years. While there appears to be indications that the TVOD indices may not be tracking the dolphin populations in a consistent manner over the whole time series, more investigation needs to be done to establish the magnitude of the problem.

References

Anon., 1998. Annual report of the Inter-American Tropical Tuna Commission. 1996. Scripps Institution of Oceanography, La Jolla, California 92037 - 1508. 306 pp.

Cochran, W.G. 1977. Sampling Techniques. John Wiley and Sons. New York, NY.

Rao, J.N.K. and C.J.F. Wu. 1988. Resampling inference with complex survey data. Journal of the American Statistical Association. 83:231-241.

Sitter, R.R. 1992a. A resampling procedure for complex survey data. Journal of the American Statistical Association. 87: 755-765.

Sitter, R.R. 1992b. Comparing three bootstrap methods for survey data. Canadian Journal of Statistics. 20: 135-154.

Smith, S.J. 1997. Bootstrap confidence intervals for groundfish trawl survey estimates of mean abundance. Canadian Journal of Fisheries and Aquatic Science. 54: 616-630.

Smith, S.J. and S. Gavaris. 1993. Improving the precision of fish abundance estimates of Eastern Scotian Shelf cod from bottom trawl surveys. North American Journal of Fisheries Management. 13: 35-47.

Thompson, S.K. and G.A.F. Seber. 1996. Adaptive Sampling. John Wiley and Sons, New York, NY.

Wade, P.R. (In press). A comparison of statistical methods for fitting population models to data. In Proceedings of the symposium on marine mammal survey and assessment methods, L. McDonald et al. (eds.). Balkeema Publishers, The Netherlands.